

This document is part of the Coordination and Support Action CRACKER. This project has received funding from the European Union's Horizon 2020 program for ICT through grant agreement no.: 645357.



Deliverable D3.2

New version of META-SHARE software

Authors: Penny Labropoulou, Miltos Deligiannis,
Juli Bakagianni, Stelios Piperidis

Dissemination Level: Public

Date: 26 December 2015

Status: Final



Grant agreement no.	645357
Project acronym	CRACKER
Project full title	Cracking the Language Barrier
Type of action	Coordination and Support Action
Coordinator	Dr. Georg Rehm (DFKI)
Start date, duration	1 January 2015, 36 months
Dissemination level	Public
Contractual date of delivery	31/12/2015
Actual date of delivery	26/12/2015
Deliverable number	D3.2
Deliverable title	New version of META-SHARE software
Type	Other (software); Report
Status and version	Final
Number of pages	9
Contributing partners	ELDA
WP leader	ATHENA RC
Task leader	ATHENA RC
Author(s)	Penny Labropoulou (ARC), Miltos Deligiannis (ARC), Juli Bakagianni (ARC), Stelios Piperidis (ARC)
Internal reviewers	Georg Rehm (DFKI)
EC project officer	Pierre-Paul Sondag
The partners in CRACKER are:	<ul style="list-style-type: none"> • Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany • Charles University in Prague (CUNI), Czech Republic • Evaluations and Language Resources Distribution Agency (ELDA), France • Fondazione Bruno Kessler (FBK), Italy • Athena Research and Innovation Center in Information, Communication and Knowledge Technologies (ATHENA RC), Greece • University of Edinburgh (UEDIN), UK • University of Sheffield (USFD), UK

For copies of reports, updates on project activities, and other CRACKER-related information, contact:

DFKI GmbH

CRACKER

Dr. Georg Rehm

Alt-Moabit 91c

D-10559 Berlin, Germany

georg.rehm@dfki.de

Phone: +49 (0)30 23895-1833

Fax: +49 (0)30 23895-1810

Copies of reports and other material can also be accessed via <http://cracker-project.eu>.

© 2015 CRACKER Consortium



Contents

1	<u>Introduction</u>	4
2	<u>META-SHARE portal</u>	4
3	<u>Licensing module updates</u>	7
4	<u>Metadata schema updates</u>	8
5	<u>Improvements in the search functionality</u>	9

1 Introduction

One of the objectives of the CRACKER project is to coordinate and support the resource sharing activities underpinning high-quality multilingual technology research and development, and to build bridges with concurrently running activities preparing multilingual digital service infrastructure(s). Resource sharing builds upon, maintains and extends the existing META-SHARE resource infrastructure (<http://www.meta-share.eu>, <http://www.meta-share.org>).

Task 3.2 aims to contribute to the smooth provision of META-SHARE services and their adaptation to users' evolving requirements and changing behaviour as well as to the changing copyright and licensing landscape. It evolves along two new versions of the META-SHARE software foreseen for M12 and M24, the first of which (M12) provides an extended and improved version of the META-SHARE platform with a streamlined and adapted data model as well as an adapted licensing toolkit. In tandem, it ensures the provision of technical support and fixing of reported bugs.

The current report documents changes and updates of the META-SHARE software and platform, and concerns the following aspects:

- portal
- licensing scheme
- schema updates and improvements
- improved search functionality.

During the next months, we will proceed to the migration of the META-SHARE nodes to the new version of the platform, implementing also mechanisms and programs that perform the conversion of the metadata records as required.

2 META-SHARE portal

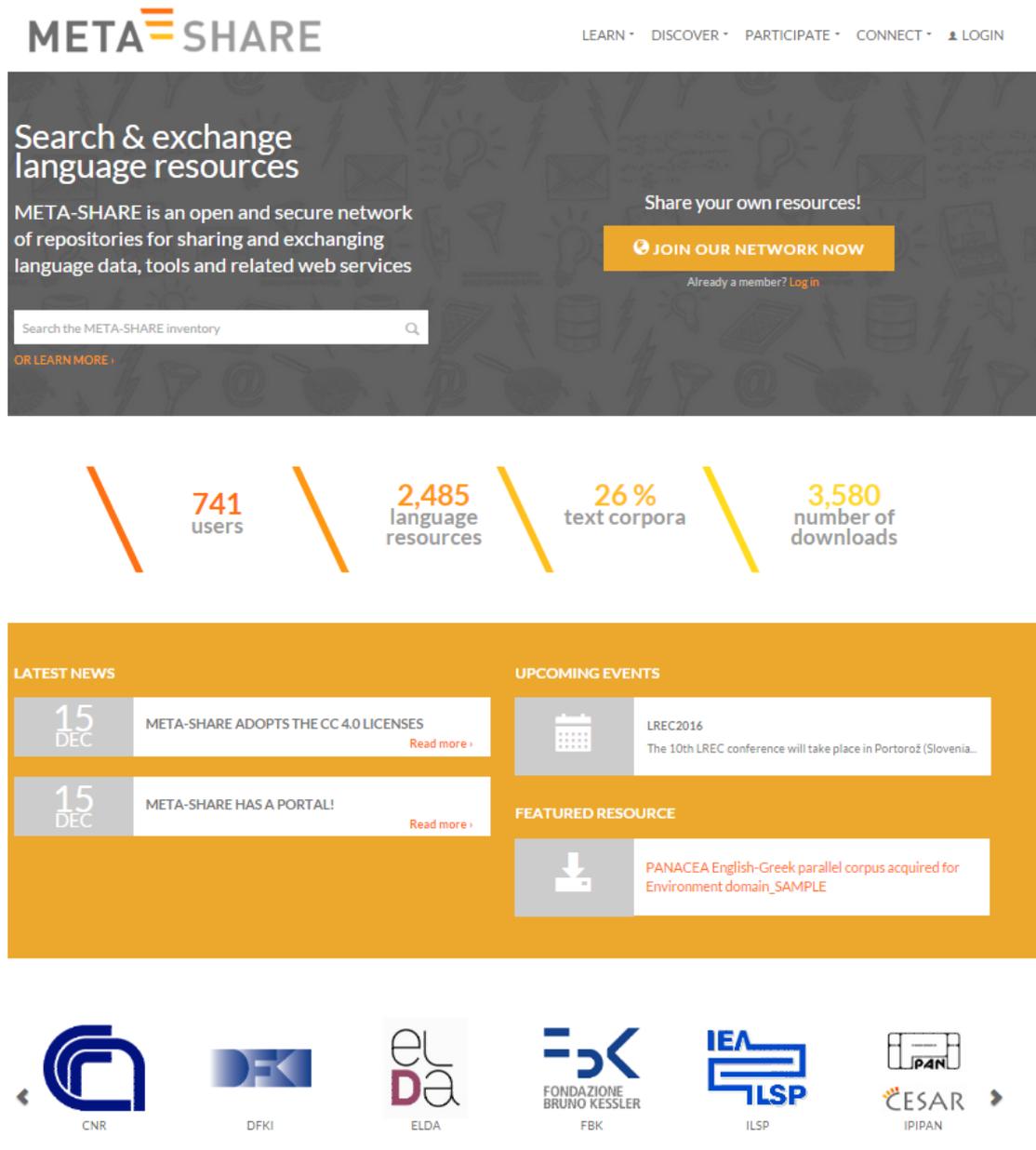
A new META-SHARE portal has been implemented which will be hosted at <http://www.meta-share.org>.

The new portal aims to facilitate the end-user when searching for language resources (LRs) in the META-SHARE infrastructure. To accomplish this, we have taken the following measures:

- re-organization of the entry page (cf. Figure 1) by removing all the complexity of presenting all managing and non-managing nodes¹; the new entry page includes a single search functionality that hits one of the META-SHARE managing nodes taking the user straight to the query results in the next screen, without having the user first select one of the nodes and then perform the search on that node.
- aggregation of all META-SHARE relevant information into one place, including the licensing scheme, user support documentation, knowledge base and forum;

¹ An explanation of differences between managing and non-managing nodes is offered at: http://metashare.dfki.de/info/#META-SHARE_Managing_Nodes.

- re-structuring of the content in order to make it easier for users to locate relevant information and navigate through the site;
- addition of support material, both in the form of new content types, such as frequently asked questions (Figure 2), events and news (Figure 3), as well as additional information on the infrastructure (Figure 4);
- a new look for the portal consistent with the familiar design scheme of META-SHARE.



META-SHARE LEARN · DISCOVER · PARTICIPATE · CONNECT · LOGIN

Search & exchange language resources

META-SHARE is an open and secure network of repositories for sharing and exchanging language data, tools and related web services

Share your own resources!

[JOIN OUR NETWORK NOW](#)

Already a member? [Log in](#)

Search the META-SHARE inventory

[OR LEARN MORE](#)

741 users

2,485 language resources

26 % text corpora

3,580 number of downloads

LATEST NEWS

15 DEC META-SHARE ADOPTS THE CC 4.0 LICENSES [Read more](#)

15 DEC META-SHARE HAS A PORTAL! [Read more](#)

UPCOMING EVENTS

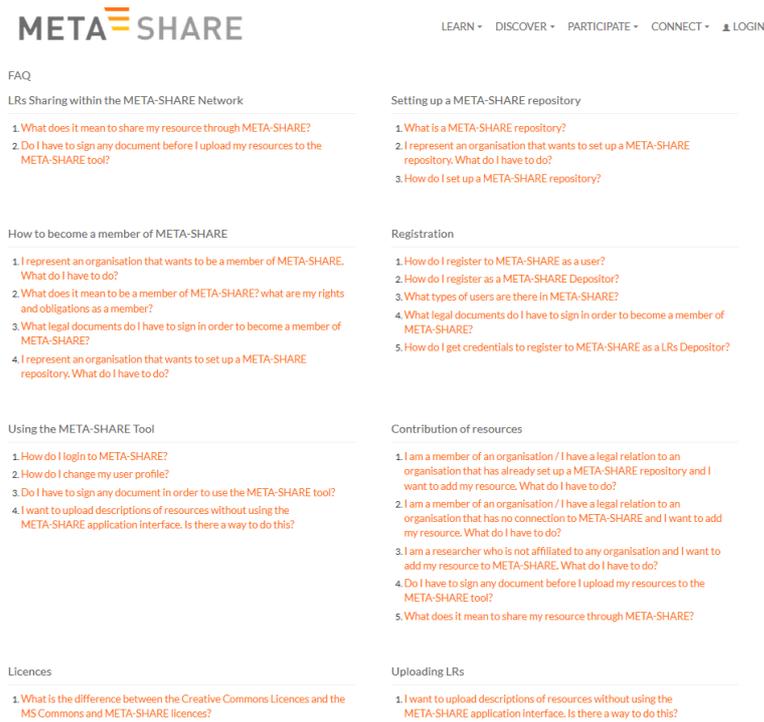
LREC2016
The 10th LREC conference will take place in Portorož (Slovenia...)

FEATURED RESOURCE

PANACEA English-Greek parallel corpus acquired for Environment domain_SAMPLE

CNR **DFKI** **ELDA** **FONDAZIONE BRUNO KESSLER FBK** **IEA ILSP** **IPAN** **CESAR**

Figure 1: Home page of META-SHARE (<http://www.meta-share.org>)



META-SHARE LEARN · DISCOVER · PARTICIPATE · CONNECT · LOGIN

FAQ

LRs Sharing within the META-SHARE Network

1. What does it mean to share my resource through META-SHARE?
2. Do I have to sign any document before I upload my resources to the META-SHARE tool?

Setting up a META-SHARE repository

1. What is a META-SHARE repository?
2. I represent an organisation that wants to set up a META-SHARE repository, What do I have to do?
3. How do I set up a META-SHARE repository?

How to become a member of META-SHARE

1. I represent an organisation that wants to be a member of META-SHARE. What do I have to do?
2. What does it mean to be a member of META-SHARE? what are my rights and obligations as a member?
3. What legal documents do I have to sign in order to become a member of META-SHARE?
4. I represent an organisation that wants to set up a META-SHARE repository, What do I have to do?

Registration

1. How do I register to META-SHARE as a user?
2. How do I register as a META-SHARE Depositor?
3. What types of users are there in META-SHARE?
4. What legal documents do I have to sign in order to become a member of META-SHARE?
5. How do I get credentials to register to META-SHARE as a LRs Depositor?

Using the META-SHARE Tool

1. How do I login to META-SHARE?
2. How do I change my user profile?
3. Do I have to sign any document in order to use the META-SHARE tool?
4. I want to upload descriptions of resources without using the META-SHARE application interface. Is there a way to do this?

Contribution of resources

1. I am a member of an organisation / I have a legal relation to an organisation that has already set up a META-SHARE repository and I want to add my resource. What do I have to do?
2. I am a member of an organisation / I have a legal relation to an organisation that has no connection to META-SHARE and I want to add my resource. What do I have to do?
3. I am a researcher who is not affiliated to any organisation and I want to add my resource to META-SHARE. What do I have to do?
4. Do I have to sign any document before I upload my resources to the META-SHARE tool?
5. What does it mean to share my resource through META-SHARE?

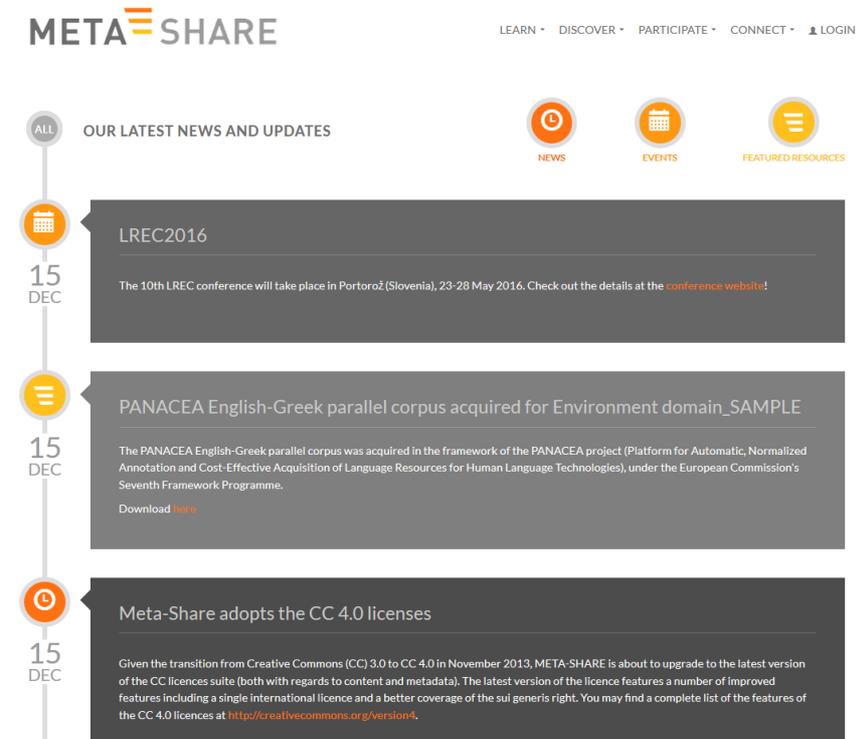
Licences

1. What is the difference between the Creative Commons Licences and the MS Commons and META-SHARE Licences?

Uploading LRs

1. I want to upload descriptions of resources without using the META-SHARE application interface. Is there a way to do this?

Figure 2: Page with frequently asked questions



META-SHARE LEARN · DISCOVER · PARTICIPATE · CONNECT · LOGIN

OUR LATEST NEWS AND UPDATES

NEWS · EVENTS · FEATURED RESOURCES

LREC2016
15 DEC

The 10th LREC conference will take place in Portorož (Slovenia), 23-28 May 2016. Check out the details at the [conference website!](#)

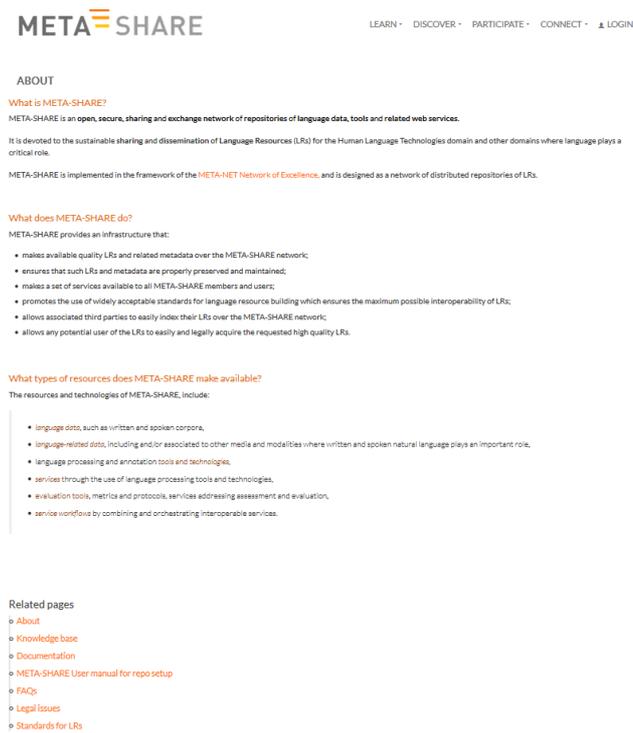
PANACEA English-Greek parallel corpus acquired for Environment domain_SAMPLE
15 DEC

The PANACEA English-Greek parallel corpus was acquired in the framework of the PANACEA project (Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies), under the European Commission's Seventh Framework Programme.
[Download here](#)

Meta-Share adopts the CC 4.0 licenses
15 DEC

Given the transition from Creative Commons (CC) 3.0 to CC 4.0 in November 2013, META-SHARE is about to upgrade to the latest version of the CC licences suite (both with regards to content and metadata). The latest version of the licence features a number of improved features including a single international licence and a better coverage of the sui generis right. You may find a complete list of the features of the CC 4.0 licences at <http://creativecommons.org/version4>.

Figure 3: Page with news and events



META-SHARE LEARN · DISCOVER · PARTICIPATE · CONNECT · LOGIN

ABOUT

What is META-SHARE?
 META-SHARE is an open, secure, sharing and exchange network of repositories of language data, tools and related web services.
 It is devoted to the sustainable sharing and dissemination of Language Resources (LRs) for the Human Language Technologies domain and other domains where language plays a critical role.
 META-SHARE is implemented in the framework of the META-NET Network of Excellence, and is designed as a network of distributed repositories of LRs.

What does META-SHARE do?
 META-SHARE provides an infrastructure that:

- makes available quality LRs and related metadata over the META-SHARE network;
- ensures that such LRs and metadata are properly preserved and maintained;
- makes a set of services available to all META-SHARE members and users;
- promotes the use of widely acceptable standards for language resource building which ensures the maximum possible interoperability of LRs;
- allows associated third parties to easily index their LRs over the META-SHARE network;
- allows any potential user of the LRs to easily and legally acquire the requested high quality LRs.

What types of resources does META-SHARE make available?
 The resources and technologies of META-SHARE, include:

- language data, such as written and spoken corpora,
- language-related data, including and/or associated to other media and modalities where written and spoken natural language plays an important role,
- language processing and annotation tools and technologies,
- services through the use of language processing tools and technologies,
- evaluation tools, metrics and protocols, services addressing assessment and evaluation,
- service workflows by combining and orchestrating interoperable services.

Related pages

- [About](#)
- [Knowledge base](#)
- [Documentation](#)
- [META-SHARE User manual for repo setup](#)
- [FAQs](#)
- [Legal issues](#)
- [Standards for LRs](#)

Figure 4: Content page of the new portal

3 Licensing module updates

An important ingredient in the META-SHARE infrastructure is the licensing module that makes clear to LR users the terms under which they can use them.

With a firm orientation towards the creation of an openness culture and the relevant ecosystem for LRs, META-SHARE has promoted the adoption of a set of standard licences, preferably supporting open re-use of resources, and more specifically Creative Commons licences for content resources and Free and Open Source-Software (FOSS) for tools and web services. In addition, a set of standard licences in the form of templates has been established to cater for the particular requirements of LR users (<http://www.meta-share.org/p/91/Licences>).

In conformance to the decision of the META-SHARE legal group² to follow the upgrades of the open licences adopted, the current deliverable also includes all the necessary amendments required for the transition of the licensing scheme to the new version of Creative Commons, namely 4.0, both with regards to content and metadata. The latest version of the licence features a number of improved features including a single international licence and a better coverage of the sui generis right.

² The META-SHARE legal group includes the following members: Prodromos Tsiavos, Stelios Piperidis, Penny Labropoulou, Khalid Choukri, Alexandros Noussias, Tasos Patrikakos, Georg Rehm, Victoria Arranz, IPR Helpdesk at ELRA.

A complete list of the features of the CC 4.0 licences is found at <http://creativecommons.org/version4>.

In order to maximise interoperability and compatibility, the META-SHARE NoRedistribution Licences v.2.0 suite has also been prepared, adopting similar terminology and wording to the CC 4.0 licences.

Finally, in this version we are suggesting to discontinue the META-SHARE Commons (MSCommons) licences. We strongly believe that MSCommons licences have been a very useful instrument that has allowed LR rights-holders to share their resources in a way that suited their needs at the time. However, since the development and release of CC 4.0 licences and as a result of our experience on the operation of different licences within the META-SHARE ecosystem, we consider that they have completed their life-cycle and can be replaced by the CC licence suite or a compatible licence such as the META-SHARE NoReDistribution (NoReD) ones.

The transition from MSCommons to CC or NoReD is of great advantage as:

- (a) the CC licences have moved to v. 4.0 and have a number of great new features especially focusing on data and databases
- (b) CC licences ensure interoperability, international legal support and easy recognition
- (c) NoReD licences allow better control of the Language Resource's distribution
- (d) NoReD allows a flexible choice of rights and restrictions
- (e) NoReD does not contribute to licence pollution or the fragmentation of the commons
- (f) the adoption of META-SHARE Commons licences has been fairly limited
- (g) we are gradually but steadily moving towards simpler, fewer, more standardised and interoperable licences.

Overall, the CC licences cover all the above conditions and, hence, we strongly support the adoption of version 4.0 of CC licences both as an upgrade from previous CC licence versions and as a replacement of the MSCommons licences.

If, further to conditions posed by CC, a data provider would like to have single-point distribution of his/her resources, i.e. distribution only from a single source, the provider may always use one of the NoReD licences.

4 Metadata schema updates

The new META-SHARE version includes changes and enhancements on the metadata schema and the browser, reflecting requests of resource providers and points of improvement that have been observed during the operation of the infrastructure. The features to be changed have been prioritized for the two deliverables; for the first version, we have opted to include mainly changes that require the least possible or easy to implement migration steps for existing metadata records:

- Correction of spelling mistakes, mainly in values of elements
- Addition of values for elements, especially if requested by metadata providers
- Change of cardinality from single to multiple for components and elements, as long as they didn't create problems for the software operation

- Addition of elements, such as the *ISLRN*³ which has been introduced as an identifier especially for language resources
- Replacement of free text values with controlled vocabularies for certain elements and, more specifically, *country*, *mimetype* and *language*-related elements (e.g. *documentLanguage*), following ISO standards and established practices
- Improvements for the licensing module, in two directions:
 - to accommodate the update of licences (i.e. update of *licence* values and texts to the CC 4.0 and MSNoReD 2.0 versions, deletion of the MSCommons values etc.),
 - to improve the presentation of licensing conditions, by adding two elements that allow LR providers to include the detailed texts of non-standard licences and terms of use/service, or point to url links thereof;
- Improvements of the appearance of metadata records:
 - a mechanism for representing the *licence* and *conditions of use* in the form of icons on the browsing page
 - highlighting of the *attribution text* and *citation paper* for the resource (if they include such information) on a prominent place in the "view record" page (Figure 5).

POETICON Multisensory and Multimedia Recordings of Everyday Interaction

▶ View resource name in all available languages

POETICON Recordings

Attribution details: The creation of this resource has been funded by the European Union, Framework Programme 7, within the framework of the POETICON project (<http://www.poeticon.eu>).

Please, acknowledge use of the resource with the following citation:

Christian Wallraven and Michael Schultze and Betty Mohler and Argiro Vatakis and Katerina Pastra, *The POETICON enacted scenario corpus - a tool for human and computational experiments on action understanding*, <http://poeticoncorpus.kyb.mpg.de/pub/fg2011.pdf>, The 9th IEEE Conference on Automatic Face and Gesture Recognition (FG '11), 2011

<http://poeticoncorpus.kyb.mpg.de/>

A corpus of language use and sensorimotor experience recordings in everyday human:human interaction, in which spontaneous language communication has been recorded along with corresponding multiview video recordings, recordings of 3D full body kinematics, and 3D tracking of objects in focus. It is a... [Read More](#)

Figure 5: View citation and attribution details in the metadata record

5 Improvements in the search functionality

The existing indexing and search functionality of META-SHARE V3.1 has been improved in the following dimensions:

- a. the Solr indexer is now endowed with a list of stopwords for English
- b. the search module is enriched with a query expansion module, which takes into account relevant synonym lists, alternative forms of frequently used query terms
- c. the ranking of the results is now based on field boosting, i.e. it prioritises fields like the *resource name*, *resource short name*, *resource description*, etc.

³ <http://www.islrn.org>